Parametric Representation of Complex Sources in Reflective Environments'

Dylan Menzies[1]

[1] *De Montfort University, UK*

Correspondence should be addressed to Dylan Menzies (`rdmg@dmu.ac.uk`)

ABSTRACT

Aspects of source directivity in reflective environments are considered, including the audible effects of directivity and how these can be reproduced. Different methods of encoding and production are presented, leading to a new approach to extend parametric encoding of reverberation, as described in the DIRAC and MPEG formats, to include the response to source directivity.

1. REVERBERATION AND SOURCE DIRECTIVITY

Natural sound sources generally do not emit a spherically symmetric sound wave. The source is said to be complex, and have directivity. The direct sound received from the source in the far-field is locally like a planewave. The direct near-field signal is more complex, but not need to be considered here (It is relevant in applications such as virtual reality where sources may need to be brought very close). By contrast the indirect sound resulting from reflections within the surrounding environment is complex, and can have a complex dependence on the source directivity. Sometimes this dependence is obvious. For example, consider a room with a long corridor leading off in one direction. If someone shouts into the corridor, the reverberation from this may be heard clearly, whereas if they shout away from it, the reverberation from the immediate room will dominate. In most cases the dependence is not so obvious, although it still has strong underlying effect on the perception of the sound scene.

Consider the image-source picture for the case of an rectangular room with simple reflections. Fig. 1 illustrates how the images of the source are oriented relative to the listener. The arrows indicate the orientation of the images in the image-source space. Each reflection originates from one direction relative to the source, and the associated sound heard by the listener will be the one radiated from the source in that direction. So, the sound making up the received reflections is a mosaic of the sound emitted

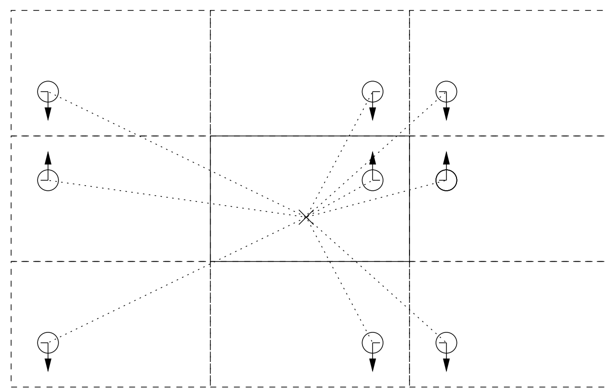

Fig. 1: Reflections calculated from directed source images. Listener at the cross.

from different directions of the source, with different time delays. This amounts to hearing a scrambled copy of the full sphere of source sound directed inwards to the listener. In a more realistic model each wall reflection is partially diffused, causing the listener to judge the source image signals to be blurred, with blurring increasing with reflection order. Blurring is realized by smoothing over a region of directions centred about the reflection. As the directed sound evolves so the received reflections capture all the changes. It is worth distinguishing between several ways in which the change can occur, and the implications for information that can be inferred by the listener.

The short time frequency distribution over direction relative to the source may be nearly constant, while the orientation of the object changes. This will cause a highly correlated change in the received reflections that could provide the listener with cues that the changes are caused by reorientation.

Objects that behave nearly as linear mechanical devices, such as bells and loudspeakers, have approximate linear responses in which each frequency radiates with a fixed directivity. Overall directivity depends on the current excitation at different frequencies.

In general sources have no fixed directivity for each frequency. This can occur because the resonant structures that produce sound do not have a fixed shape. This is particularly true for instance in the human voice where the shape of the vocal tract changes, and woodwind instruments where toneholes are used to dynamically reconfigure resonances. The listener experiences a change of directivity for the same pitch produced in different ways. This can provide a greater appreciation of the state of the instrument, and mode of expression.

2. RECORDING

Spatial recording techniques, such as stereo and Ambisonics, encode direction information about real reflections, and so changes associated with source directivity are also captured. Recordings made in this way can have rich and dynamic reflective components. Many other audio recordings are produced from dry recordings of instruments with artificial reverberation added afterwards. Reverberation processors usually have one input, and do not model the source directivity, resulting in a more static, less responsive reverberation.

The following experiment demonstrates the use of source directivity, and how it can be applied in a studio environment. The source is recorded using two cardioid microphones each pointing towards the centre of the source but from different angles. For example, a woodwind instrument radiates with significant directivity across a range of angles depending on the finger configuration. Two microphones angled 90 degrees about the front together capture much of the sound while also being distinct. These are fed into two reverberators with slightly different settings for early reflections and delay offsets, see Fig. 2. It is apparent that the reverberation becomes more dynamic, involving, and lifelike. Note this this process is distinct from processing a stereo recording, which is an approximation of the freefield, whereas here an approximation to the source multipole field is being recorded.

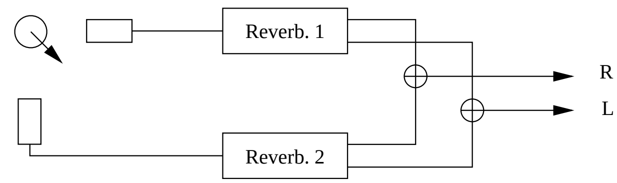

Fig. 2: Recording of a directed source with two mics and processing with two stereo reverberators.

3. LINEAR SOURCE REPRESENTATION

We wish now to elaborate the process just described to allow for a more detailed description of the complex or directed source and the response of the environment. Acoustic theory allows us to define a general source completely by a signal defined in each direction from a centre chosen in the source, or equivalently a complex function of frequency in each direction. The centre can be chosen freely, although there is usually an optimum position which minimizes the complexity of the description. The set of directions can equally be represented as points on the surface of a sphere. This can be decomposed into signals multiplying spherical harmonic functions of direction. One advantage of this is that a finite number of signals is required to represent the source up to a given resolution. Also, this choice of direction basis happens to extend to a full set of basis functions for the wave equation in space, including the radial extent. These functions are sometimes called multipoles. The detailed form of the radial dependence is only needed for studying the near-field of the source, which is not considered here.

The environment filters the source signal into a reverberant signal received by the listener. This is modelled completely as a Fourier-Bessel expansion (FBE), also known as High Order Ambisonic (HOA) encoding, which is the free-field analogue of the multipole source expressed in spherical harmonics. Assuming linearity of the environment, which is usually nearly true, the FBE is a linear function of the multipole signals, order by order, [1]. This function can be measured by driving directed sources with sine sweeps and deconvolving the sweep from the directed microphone listener signals [2]. The room response can then be applied to any directed source signal to generate the detailed reverberant response of the room at the listener Angular resolution of the listener signal, which includes the whole environment, is more important than angular resolution of the source, and there is no point having a higher resolution for the source than the receiver.

4. PARAMETRIC REPRESENTATION

The linear response matrix described above is comprehensive and well defined, but it is also complex and potentially wasteful from a psychoacoustic perspective. If we consider the initial room response from an impulse emitted by a complex source, it consists of a succession of reflections arriving from different directions and separated in time. Each of these reflections originates from one direction from the source, as shown in Fig. 1. Diffuse reflection will cause a blurring of this direction, so that the received signal is an average over a region of the source. As reflections overlap in time, there will be simultaneous contributions from different directions of the source, again blurred.

Parametric encoding of reverberation for a monopole source was introduced in SIRR, and developed into the DIRAC spatial codec, and related MPEG codecs, [3, 4, 5]. The psychoacoustic principle used is that in each frequency band, the spatial properties of received sound are represented well by a single direction $\Theta(t, f)$ and proportion of energy that is diffuse $\Psi(t, f)$. In reconstruction a direct signal and a omnidirectional decorrelated signal are generated in each band.

This approach can be extended to a directed source by adding parametric information describing the origin of each received time-frequency tile on the source. For an isolated reflection this will be centred on a well defined direction on the source $\Theta_S(t, f)$ with an angular spread $\Omega_S(t, f)$ caused by diffuse reflection, see Fig. 4. A limitation with this parametrization is that for each tile, varying the source directivity will not effect the reflection direction distribution, only its overall gain. This is nearly

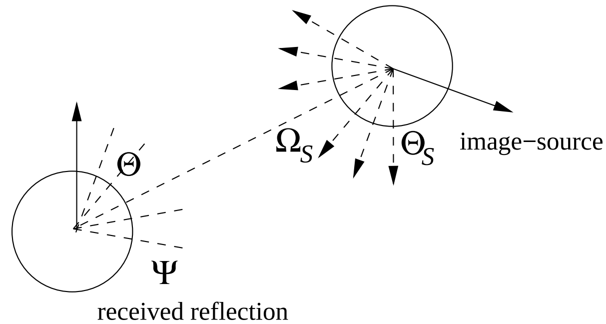

Fig. 3: Extended parametrization to capture source directivity.

true for a single reflection in a tile. In reconstruction the region of the source described by $\Omega_S(t, f)$ is averaged to find the signal that is then processed using the SIRR technique above. When multiple reflections coincide in a tile the reflection distribution can vary greatly depending on the source directivity. Potentially these distributions could be encoded as extra parameters. However, reflection overlap rapidly increases to the point where the reflection signal is mostly diffuse and there is little use in having many separate reflection distributions. The early reflections are most effective at communicating information about the source directivity.when reflections overlap, varying the source directivity will not cause the reflected directivity.

The source can be represented by the previous linear method, or using a parametric method. The parametric approach with a single direction is less useful for more complex objects, as multiple maxima and minima of sound intensity over direction are valuable features. In a 1st order linear representation, single maxima can be represented, in 2nd order two maxima and so on. Possibly a feature based representation, such as a gaussian mixture model could be an efficient approach for more complex sources.

The new parametric response variables, $\Theta_S(t, f)$ and $\Omega_S(t, f)$ can be calculated by tracing the origin of each response tile. For recorded or modeled reverberant sound this can be achieved by using multiple measurements with different source directions or harmonics. For each tile the relative received magnitude resulting from each direction from the source can be mapped as a distribution function over the sphere. $\Theta_S(t, f)$ and $\Omega_S(t, f)$ are taken respectively as the centre and spread of this distribution.

5. SUMMARY

Modeling room response to source directivity is useful for useful for creating a more natural impression of sound sources, that may also contain important acoustic cues. Simple studio techniques can be used to create an approximate effect. This can be extended rigorously to an accurate representation with a filter matrix. A more compact representation is formed by adding information to a parametric representation of the response. The extra information codes the directions on the source that contribute to each time frequency tile.

6. REFERENCES

[1] D. Menzies and M. Al-Akaidi. Ambisonic synthesis of complex sources. *Jounal of the Audio Engineering Society*, 55(10):864–876, October 2007.

[2] A. Farina. Measuring impulse responses containing complete spatial measuring impulse responses. In *Proc. 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan*, 2006.

[3] J Merimaa and V Pulkki. Spatial impulse response rendering I: Analysis and synthesis. *JOURNAL OF THE AUDIO ENGINEERING SOCIETY*, 53(12):1115–1127, DEC 2005.

[4] Ville Pulkki. Spatial sound reproduction with directional audio coding. *JOURNAL OF THE AUDIO ENGINEERING SOCIETY*, 55(6):503–516, JUN 2007.

[5] Juha Vilkamo, Tapio Lokki, and Ville Pulkki. Directional Audio Coding: Virtual Microphone-Based Synthesis and Subjective Evaluation. *JOURNAL OF THE AUDIO ENGINEERING SOCIETY*, 57(9):709–724, SEP 2009.